Mapping irrigated rice using MSI/Sentinel-2 time series of vegetation indices and Random Forest

Juliana de Abreu Araújo[1], Allan Henrique Lima Freire[1], Ricardo Dalagnol [1], Lênio Soares Galvão[1]

[1] Earth Observation and Geoinformatics Division - National Institute for Space Research (INPE)
Postal Code 515 - 12227-010 - São José dos Campos – SP – Brazil

{juliana.araujo, allan.freire, ricardo.silva, lenio.galvao}@inpe.br

Abstract. *The State of Rio Grande do Sul is the largest producer of irrigated rice in Brazil. The goal of this study was to use a time series of MSI/Sentinel-2 vegetation indices (VIs), such as Normalized Difference Vegetation Index – NDVI; and Normalized Difference Water Index – NDWI, to extract metrics of irrigated rice cultivation in the municipality of Uruguaiana-RS and map this crop in the 2019/2020 period using Random Forest (RF). Our methodology generated maps with a systematic and cost-efficient benefit for mapping rice between the 2019/2020 period (with a 96,5% of overall classification accuracy), showing consistency with those of available maps from official institutions, and for predicting production in the 2020/2021 period.*

1. Introduction

The monitoring of agricultural crops is important for economic development, food security, and environmental conservation [Laborte, 2017]. The applications range from forecasting production volume to government management of agricultural stocks, such as export/import stimulus, availability of rural credit, and support for agricultural insurance. The monitoring contributes to assist in the decision making for buying, selling, distribution, and national supplying of the agents in this chain. In many places in the world, especially in underdeveloped countries, official data regarding crop areas is done based on farmers' reports or field visits. This process can be bureaucratic because of the time that it takes to organize data and field visits [Frolking, 2002]. In this circumstance, remote sensing images allow a systematic and comprehensive analysis of a given areas quickly and at low cost [Castro, 2020].

Different digital image processing methods have been developed and are required to monitor crops. The dissemination of medium resolution products in multi-sensor remote sensing, such as MSI/Sentinel-2 series (10-m spatial resolution), have boosted precision agriculture activities and the advancement in research of crop mapping and determination of crop areas. This is because of its practicality and ability to collect information from land covers on the Earth's surface with improved spectral and spatial resolutions.

An example of a plantation of great relevance for Brazil is the irrigated rice. It is inserted in the conditions of constant updates for crop mapping research because of its great productive representativeness that aims to supply the country's domestic supply, in addition to the contribution of crops to the country's economy, since according to the

National Supply Company (CONAB, 2020), 76% of the production of all irrigated rice in Mercosur comes from Brazil. However, the extensive planted area is an obstacle to faster mapping that can be circumvented with the use of remote sensors of medium resolution and with a dense temporal coverage for the construction of systematized time series. This allows an effective assessment of the phenological structure of the crop and its behavior during the entire cycle of development from soil preparation up to harvest.

The main goal of this study was to map irrigated rice in a small area of south Brazil. For this purpose, we used times series of two vegetation indices (NDWI and NDVI), calculated from MSI/Sentinel-2 satellite data, to extract metrics of irrigated rice cultivation to use as input data to a Random Forest (RF) model to process and identify a spatialized classification of rice cultivation in Uruguaiana. Compared to the Landsat instruments, the use of a 5-day temporal resolution and of a 10-m spatial resolution for some MSI bands increases the change of obtaining cloud-free data and mapping small fields of irrigated rice.

2. Material and Methods

The study area has 5,702 km² (Figure 1) and is located in the Uruguaiana municipality in the State of Rio Grande do Sul. This area is known by the historical use of irrigation in Brazil, where rice farming has been active since the beginning of the last century.

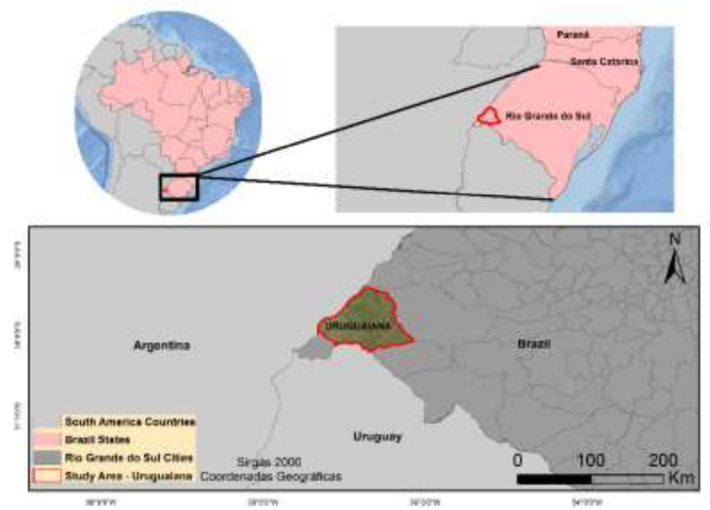

Figure 1. Study Area.

The first step of the methodology was to obtain the satellite time series of the Sentinel 2A and 2B images (Figure 2). Every scene of surface reflectance image was collected from August 2019 to June 2020 and from August 2020 to June 2021, were used to calculate the Normalized Difference Vegetation Index (NDVI) [Rouse/1974] and the Normalized Difference Water Index (NDWI) [Gao/1996] on the Google Earth Engine (GEE) computing platform, then the mosaic of the scenes was performed to contemplate the entire study area (four images to compose a scene). In order to fill the gaps caused by the cloud mask, a cubic interpolation was applied to the time series. In addition, a Whittaker smoother was applied [Whittaker/1923] in order to remove noise from the time series.

The crop period was chosen based on the CONAB (2019) agricultural calendar, extending into one month before planting to understand the flooding period of the rice crop plots. In spite of the adequate spatial resolution of the MSI/Sentinel-2 for mapping crops, we decide to resample the data to 100-m pixel size to optimize the processing time to extract the metrics.

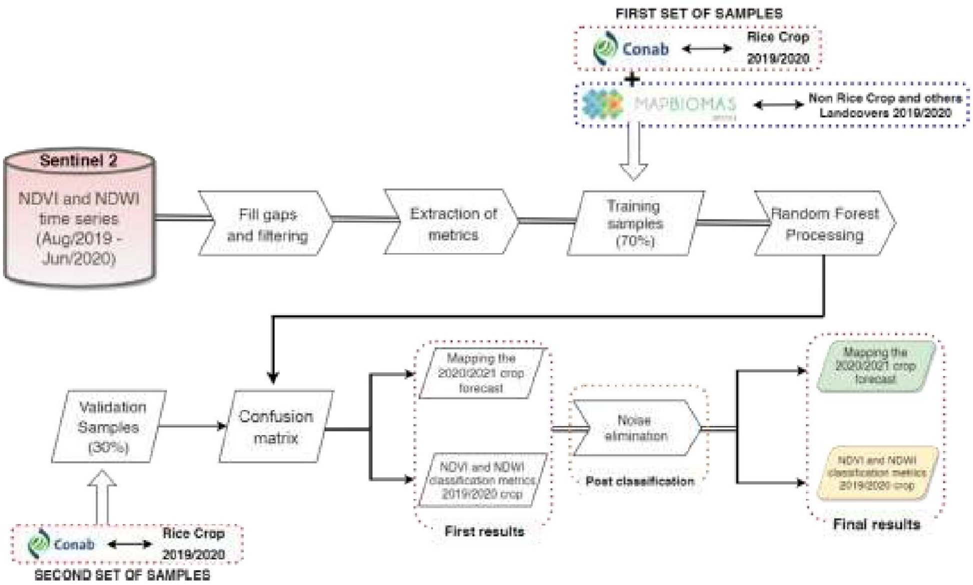

Figure 2. Flowchart with the process steps.

After gap filling and filtering, we extracted basic and polars metrics from the time series [Korting/2012]. The basic metrics obtained were: derivative absolute mean, mean, minimum value, maximum value, standard deviation, and amplitude. The polar metrics represent the time series projected in polar coordinates [Bendini/2020], thus it is possible to calculate the area of each quadrant, as described by Korting (2012). The metrics describe the objects' properties and help to select specific characteristics and behaviors that allow distinguish between the different objects present in a scene [Korting/2012].

The RF machine learning algorithm was used for classification of rice crops, considered as a supervised classification algorithm. The basis of RF is tied to a combination of decision models, also called decision trees, which are responsible for improving the accuracy of the classification [Breiman/2001]. Because it has this supervised feature, it was necessary to create a set of refined and adjusted training samples between two classes of rice and non-rice crops.

The samples used in the training and testing of the RF model were obtained from the irrigated rice mapping performed by CONAB for the 2019/2020 growing season. The agriculture mask was obtained from MapBiomas (2020). Rice points samples were randomly selected within the polygons of the CONAB rice mapping, while non-rice points samples were randomly obtained from the polygons of the discounted agriculture mask of the CONAB rice mapping. In order to reduce spectral mixture of samples, near a 100-meter buffer was applied before extracting the samples.

A total of 1000 points were extracted (500 for each class), which were then carefully evaluated to verify the correct positioning of the sample in the corresponding area, that is, away from edges or pixels without accurate information about mapped land covers (rice and non-rice).

The model accuracy was evaluated considering the confusion matrix and the determination of metrics such as the overall classification accuracy, recall and precision. We compared RF models using different input data: (1) basic metrics, (2) polar metrics, and (3) the combination of basic and polar metrics.

3. Results and Discussion

The rice and non-rice samples showed remarkably different patterns in the time series of NDVI and NDWI (Figure 3). At the time before the maximum vegetative vigor of the grain (Feb/2020), the NDWI had a stable behavior, especially during land preparation and flooded stages for rice development. This curve seeks its minimum values when the NDVI has its peak, demonstrating the maximum vegetative vigor of the crop and the absence of water background influence in this period. This typically happens between January and February, when the crop is about to be harvested. In the non-rice temporal profiles, we noticed a non-standardization of the behavior of the NDVI and NDWI curves because of the mixture of targets present in this class.

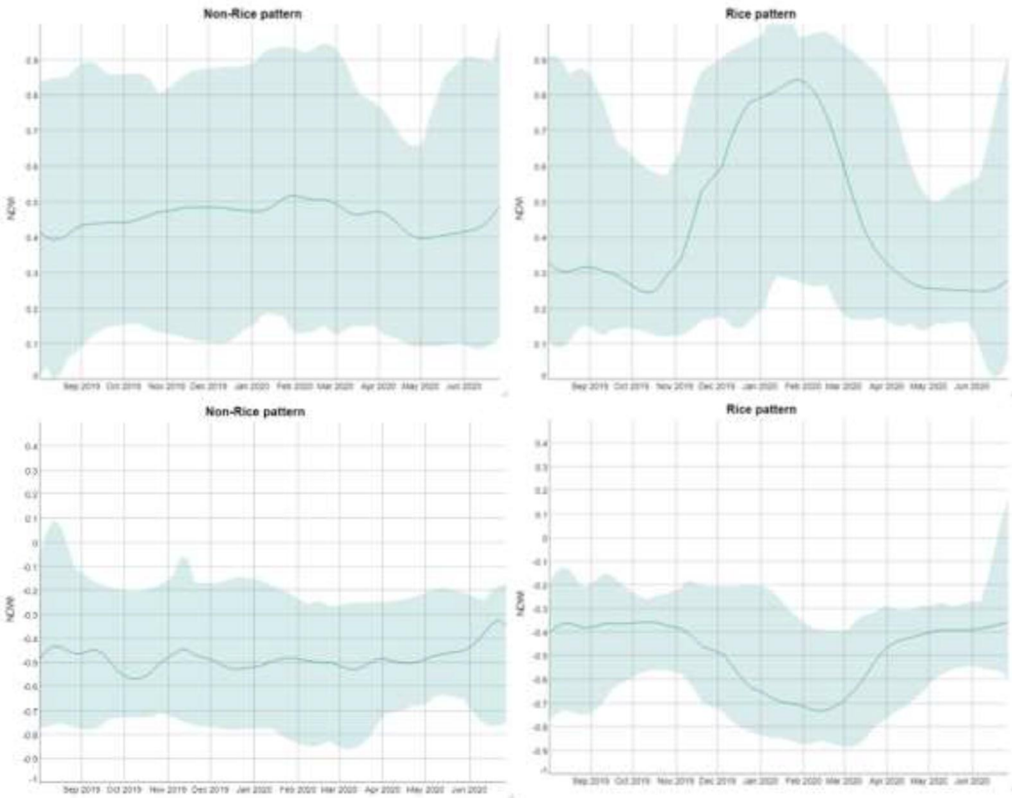

Figure 3. Temporal patterns of vegetation indices (NDVI and NDWI) for selected samples, for 2019/2020 time series.

Table 1 and Table 2 show the performance of the NDVI and NDWI derived metrics to separate rice fields from non-rice areas, respectively. Results refer to the models trained and tested with the basic, polar, basic + polar metrics. Because of the large presence of rice (defined behavior) in the region and the continuous sampling throughout the study area, the overall accuracy was quite promising (between 95% and 98%). This result indicates separability between the classes.

Table 1. Confusion matrix for crop classification using NDVI and the RF model during the 2019/2020 growing season.

			Rice	Non-Rice	Total	Accuracy	Recall	Precision	F1	Estimated area
NDVI CROP 2019/2020	Basics	Rice	130	8	138	95%	94%	94%	0,94	**79,55 mil Ha**
		Non-Rice	8	154	162					
		Total	138	162	300					
	Polar	Rice	132	6	138	96%	95%	96%	0,95	**81,44 mil Ha**
		Non-Rice	7	155	162					
		Total	139	161	300					
	Basics + Polar	Rice	132	6	138	97%	98%	96%	0,97	**79,42 mil Ha**
		Non-Rice	3	159	162					
		Total	135	165	300					

Table 2. Confusion matrix for crop classification using NDWI and the RF model during the 2019/2020 growing season.

			Rice	Non-Rice	Total	Accuracy	Recall	Precision	F1	Estimated area
NDWI CROP 2019/2020	Basics	Rice	145	8	153	96%	97%	95%	0,96	**77,12 mil Ha**
		Non-Rice	5	142	147					
		Total	150	150	300					
	Polar	Rice	155	3	158	98%	98%	98%	0,98	**81,34 mil Ha**
		Non-Rice	3	139	142					
		Total	158	142	300					
	Basics + Polar	Rice	146	7	153	97%	98%	95%	0,97	**76,68 mil Ha**
		Non-Rice	3	144	147					
		Total	149	151	300					

Tables 1 and 2 also show the 3 models developed for the individual metrics. Again, a high classification accuracy (95%) was observed, especially for the NDVI Basic + Polars metrics that had mapped areas of rice similar to the reference values provided by CONAB (79.7 thousand ha) (Table 1).

Sample location, data pre-processing and time series smoothing contributed to classification accuracy of the RF classifier. Because of careful sampling and positioning of the samples, a step that demanded time in the work, the accuracy of the models was high (above 95%). This was reflected in the final classification of the rice areas that were well identified and with few classification errors, as deduced from the confusion matrix tables.

After removing the noise present in the classifications in all metrics, visually, the classification of the areas of the 2019/2020 crop was spatially well defined and in accordance with the manual classification made by CONAB. The classification result using basic + polar metrics for NDVI, and polar metrics for NDWI, are presented in Figures 4 and 5, respectively.

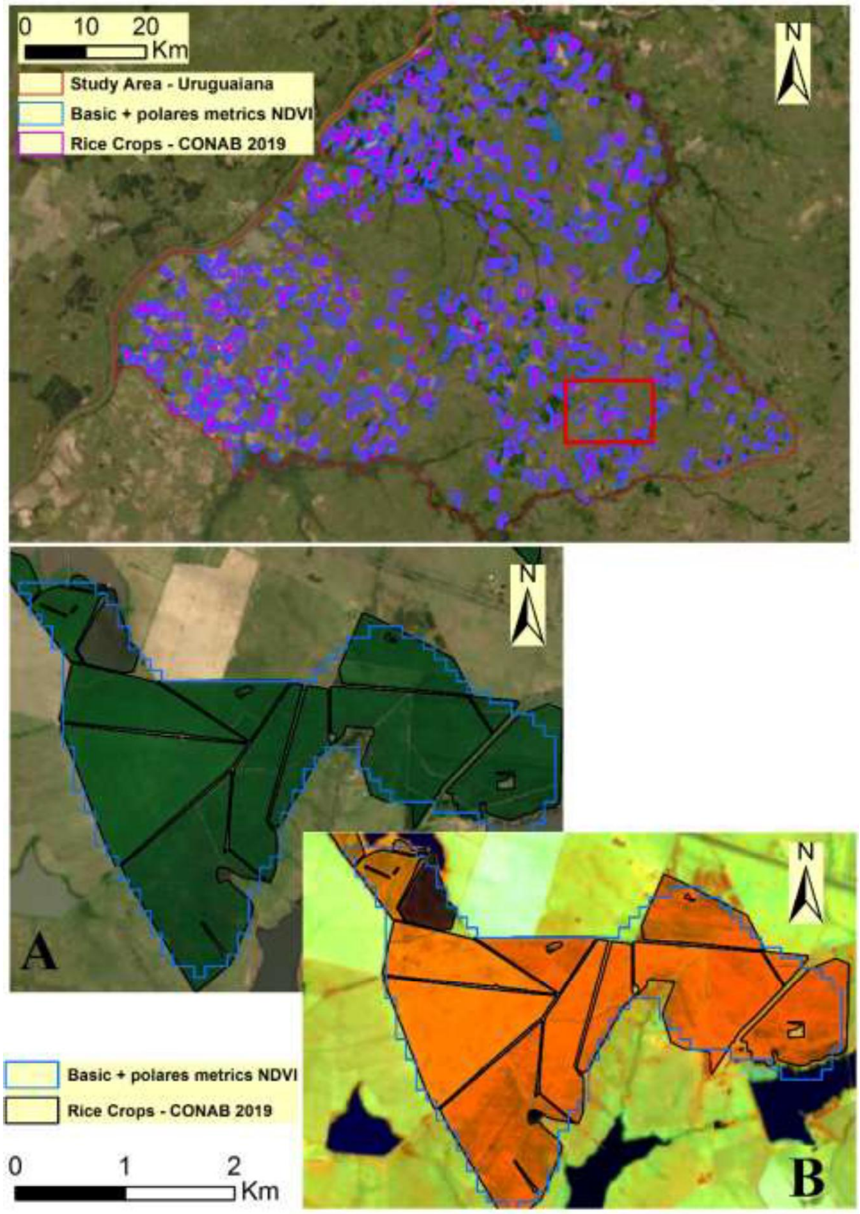

Figure 4. Classification results using basic and polar metrics extracted from NDVI time series for the 2019/2020 growing season.

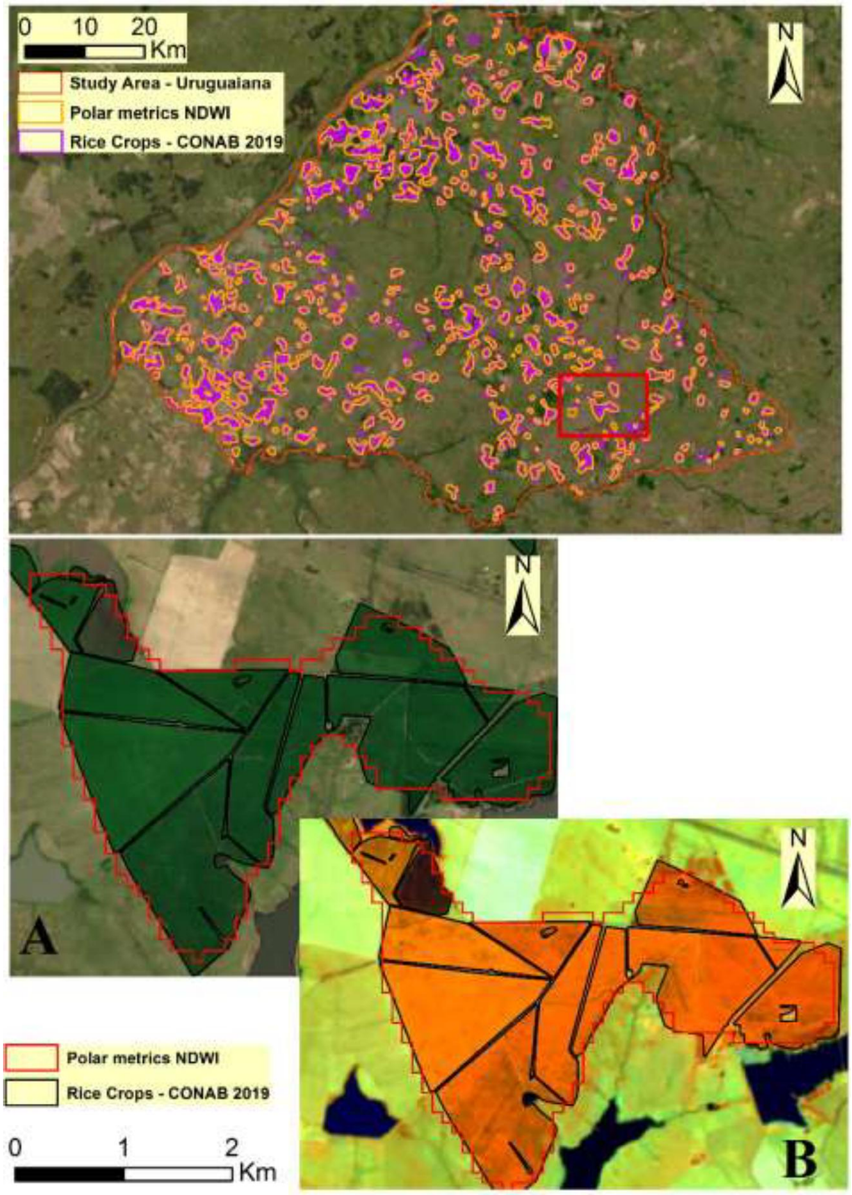

Figure 5. Classification results using polar metrics extracted from NDWI time series for the 2019/2020 growing season.

From these figures, we can observe consistency between the RF and CONAB maps, even considering the constraint of data resampling to 100-m pixel size in our data set.

We also performed crop prediction for currently irrigated rice areas for the 2020/2021 period (Figure 6). Once again, the algorithm behaved promisingly when correctly classifying crop areas using the polar + basic metrics RF model. The total estimated area was 5% larger in relation to the area of the last 2019/2020 harvest period, with about 83.65 thousand ha, according to the estimated classified area assigned by the NDVI metric bounded by the RF algorithm for the 2020/2021 crop.

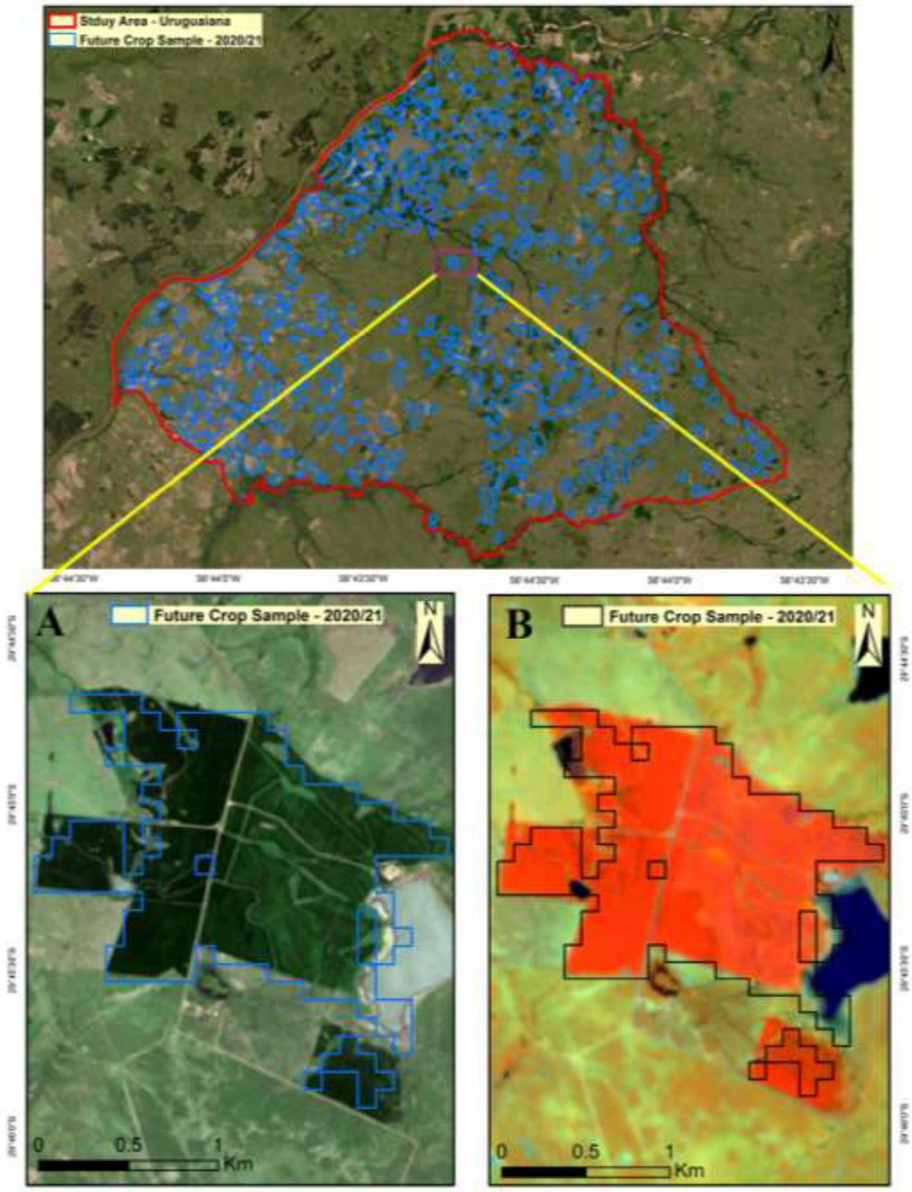

Figure 6. Crop prediction results using polar + basic metrics model extracted from the NDVI time series for the 2020/2021 period.

4. Conclusions

In the present work, we sought to identify irrigated rice in the municipality of Uruguaiana using image processing techniques and machine learning. The combined use of basic and polar metrics from MSI/Sentinel-2 time series of VIs and the RF model allowed classification of irrigated rice areas with accuracy above 95% for the 2019/2020 growing season. The accuracy of the models and the estimated areas showed slight differences across VIs. For the NDVI time series, the best accuracy (97%) was observed in the basic metrics + polar model, in which the area was underestimated by 1%. For the NDWI time series, the best accuracy (98%) was observed in the polar metrics model, in which the area was overestimated by 2%.

The methodology was promising for the classification of irrigated rice areas and for crop prediction. For future work, it is recommended to extrapolate and test the application of the model to other regions and to use the images with the original spatial resolution of the MSI/Sentinel-2 (10-m) to improve the current maps.

5. References

Companhia Nacional de Abastecimento (CONAB). (2020) Acompanhamento da Safra Brasileira: Grãos, Quarto Levantamento —janeiro/2020, v.7, p.1-104.

Laborte, A.G., Gutierrez M.A., Balanza J.G., Saito K., Zwart S.J., Boschetti M., Murty M.V.R., Villano L., Aunario J.K., Reinke R., and et al. (2017) Riceatlas, a spatial database of global rice calendars and production. *In Scientific Data*, v.4, p.1-10.

Frolking, S., Qiu J., Boles S., Xiao X., Liu J., Zhuang Y., Li C., and X Qin. (2002) Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in china. *In Global Biogeochemical Cycles - GLOBAL BIOGEOCHEM CYCLE.*

Castro Filho, H. C. Carvalho O.A.J Carvalho O.L.F. de Bem P.P. Moura R.S. Alburquerque A.O. Silva C.R. Ferreira P.H. G. Guimarães R.F. and R.A.T Gomes. (2020) Rice crop detection using lstm, bi-lstm, and machine learning models from sentinel-1 time series. *In Journal*, v.12, p.1-25.

Jr. Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. (1974) Monitoring Vegetation Systems in the Great Plains with Erts.

Gao. B.C. (1996) NDWI a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, v.58, p.1- 9.

Companhia Nacional de Abastecimento (CONAB). (2019) Calendário de Plantio e Colheita de Grãos no Brasil, p.75.

Whittaker E.T. On a new method of graduation". (1923) *In proceedings of the Edinburgh Mathematical Society*, v.41, p.1-12.

Korting. T.S. (2012) Geodma: a toolbox integrating data mining with object-based and multi-temporal analysis of satellite remotely sensed imagery, p.1-123.

Bendini. H.N., Leila M. G. Fonseca, Anderson R. Soares, Philippe Rufin, Marcel Schwieder, Marcos A. Rodrigues, Raian V. Maretto, Thales S. Korting, Pedro J. Leitão, Ieda D. A. Sanches, and Patrick Hostert. (2020) Applying a phenological object-based image analysis (phenobia) for agricultural land classification: A study case in the brazilian cerrado. *In proceedings…* pages 1078–1081.

Breiman. L. (2001) Random forests. *In Machine Learning*, v.45, p.1-33.

MapBiomas. (2020) Projeto de mapeamento anual do uso e cobertura da terra no brasil, https://plataforma.brasil.mapbiomas.org/, July.